Chapter 1
Convex analysis

Convex Optimization for Computer Vision and Machine Learning
WS 2017

Virginia Estellers, Emanuel Laude
Computer Vision Group
Faculty of Informatics
Technical University of Munich

Mathematical basics

Open, Closed, and Compacts Sets

Definitions

A set $C \subset \mathbb{R}^n$ is **open** if for all $x \in C$ there is $\epsilon > 0$ s.t. the ball of radius ϵ around x, $B(x, \epsilon)$, is contained in C:$B(x, \epsilon) \subset C$
A set $C \subset \mathbb{R}^n$ is **closed** if its complement is open
A set is closed if and only if it contains all its limit points.
The **closure** $\overline{C}$ of a set C is

$$\overline{C} = \{x \mid \text{ there is a sequence } (x_n)_n \subset C \text{ s.t. } \lim_{n\to\infty} x_n = x\}$$

The **interior** $\mathring{C}$ of C is $\{x \in C \mid \text{ there is } \epsilon > 0 \text{ s.t. } B(x, \epsilon) \subset C\}$
A set $C \subset \mathbb{R}^n$ is **compact** if it is closed and bounded

Bolzano-Weierstrass Theorem

Let $(x_n)_{n\in\mathbb{N}} \subset C$ be a sequence in the compact set C. Then there is a convergent subsequence (x_{n_k}) with $\lim_{k\to\infty} x_{n_k} = \hat{x} \in C$

Continuity

Definition: Lower semi-continuity

A function $E : \mathbb{R}^n \to \mathbb{R}$ is **lower semi-continuous** (l.s.c.), if for all u it holds that

$$\liminf_{v \to u} E(v) \geq E(u)$$

Definition: Lipschitz Continuity

A function $f : C \subset \mathbb{R}^n \to \mathbb{R}^m$ is **Lipschitz continuous** with *Lipschitz constant* L if for all $x, y \in C$

$$\|f(x) - f(y)\|_2 \leq L\|x - y\|_2$$

A function $f : C \subset \mathbb{R}^n \to \mathbb{R}^m$ is **locally Lipschitz continuous** if for every $x \in C$ there exists $\epsilon > 0$ such that $f_{|B(\epsilon,x)}$ is Lipschitz continuous

Convexity

Convex energy minimization problems

A convex energy minimization problem is

$$\hat{u} \in \arg\min_{u \in C} E(u),$$

where $C \subset \mathbb{R}^n$ convex set, $E : C \to \mathbb{R}$ convex function.

1. What is a convex set?

Definition

A set $C \subset \mathbb{R}^n$ is called convex, if

$$\alpha x + (1-\alpha)y \in C, \quad \forall x, y \in C,\ \forall \alpha \in [0,1].$$

Convex set

The following operations preserve the convexity of a set

- Intersection
- Minkowski sum
- Closure
- Interior
- Linear Transformation

The union of convex sets is not convex in general.

Polyhedral sets are always convex, cones not necessarily.

Convex energy minimization problems

$$\hat{u} \in \arg\min_{u \in C} E(u),$$

where $C \subset \mathbb{R}^n$ convex set, $E : C \to \mathbb{R}$ convex function.

1. What is a convex set? We know this now

2. What is a convex function?

Definition: Convex Function

Function $E : C \to \mathbb{R}$ is **convex** if C is a convex set and for all $u, v \in C$ and all $\theta \in [0, 1]$ it holds that

$$E(\theta u + (1 - \theta)v) \leq \theta E(u) + (1 - \theta)E(v)$$

E is **strictly convex** if the inequality is strict for all $\theta \in (0, 1)$, $v \neq u$

Convex functions

The following operations do preserve the convexity of a function

- Non-negative weighted sum
- Composition with an affine function
- Pointwise maximum and supremum

Useful

$E : C \subset \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for any $x, y \in C$,
$F(t) = E(x + ty)$ is convex for all $t \in \mathbb{R}$ such that $x + ty \in C$

The sum of a convex function and a strictly convex function is strictly convex.

Extended Real-valued Functions

Given a convex set C and a convex function $E : C \to \mathbb{R}$

$$\hat{u} \in \arg\min_{u \in C} E(u), \tag{1}$$

can be formulated as an unconstrained minimization by "introducing" the constraint $u \in C$ into the energy function E.

Formally, (1) is written as

$$\hat{u} \in \arg\min_{u \in \mathbb{R}^n} \tilde{E}(u),$$

in terms of the **extended real-valued function** $\tilde{E}$

$$\begin{aligned} \tilde{E} :& \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\} \\ & u \mapsto \tilde{E}(u) = \begin{cases} E(u) & \text{if } u \in C, \\ \infty & \text{else.} \end{cases} \end{aligned}$$

Revisiting the definition of convex functions

The function $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ is **convex** if

- its domain $\text{dom}(E) := \{u \in \mathbb{R}^n \mid E(u) < \infty\}$ is a convex set.
- For all $u, v \in \text{dom}(E)$ and all $\theta \in [0, 1]$ it holds that

$$E(\theta u + (1 - \theta)v) \leq \theta E(u) + (1 - \theta)E(v)$$

E is **strictly convex** if the inequality is strict for all $\theta \in (0, 1)$, $v \neq u$

Defintion: Proper Function

Function $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ is **proper** if its domain is not empty

Connection between Convex Sets and Functions

Defintion: Epigraph

Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper function, its epigraph epi(E) is

$$\text{epi}(E) := \{(u, \alpha) \mid E(u) \leq \alpha\}$$

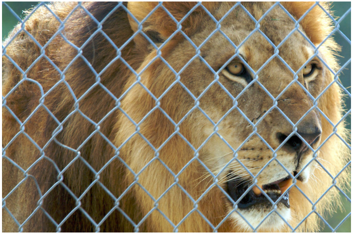
Theorem

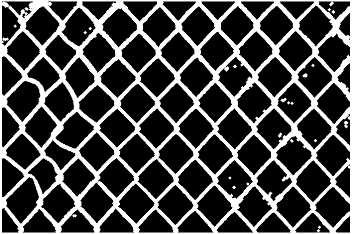
A proper function $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if and only if its epigraph is convex

Proof: Board.

First example of an imaging problem: Inpainting

Example: Inpainting



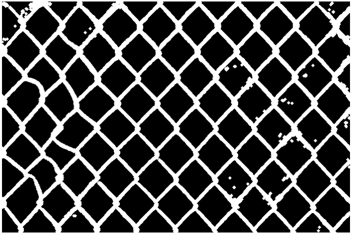


$$\min_{u \in \mathbb{R}^{n \times m}} \sum_{i,j} (u_{i,j} - u_{i-1,j})^2 + (u_{i,j} - u_{i,j-1})^2 \quad \text{s.t. } u_{i,j} = f_{i,j} \ \forall (i,j) \in I$$

with index set I of pixels to keep and suitable boundary conditions.

First example of an imaging problem: Inpainting

Example: Inpainting



$$\min_{u \in \mathbb{R}^{n \times m}} \sum_{i,j} (u_{i,j} - u_{i-1,j})^2 + (u_{i,j} - u_{i,j-1})^2 \quad \text{s.t. } u_{i,j} = f_{i,j} \; \forall (i,j) \in I$$

with index set I of pixels to keep and suitable boundary conditions.

Properties of convex functions

Theorem

Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex. Any local minimum of E is global.

Theorem: Monotonicity of the gradient

Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, convex and differentiable at $u \in \text{dom}(E)$.

$$E(v) - E(u) - \langle \nabla E(u), v - u \rangle \geq 0 \qquad \forall v \in \mathbb{R}^n$$

Proofs: Board.

Alternative Definition

Function $E : C \to \mathbb{R}$ is convex if and only if C is convex and for all $u, v \in C$, $\beta \geq 0$ such that $u + \beta(u - v) \in C$ it holds that

$$E(u + \beta(u - v)) \geq E(u) + \beta(E(u) - E(v))$$

Properties of convex functions

The behavior of convex functions at the boundary of their domain can be out of control if unless they are closed.

Definition: Closed convex function

A convex function is closed if its epigraph is closed.

For instance:

$$E(x, y) = \begin{cases} 0 & \text{if } x^2 + y^2 < 1 \\ \phi(x, y) & \text{if } x^2 + y^2 = 1 \end{cases}$$

- $\mathrm{dom} E = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 \leq 1\}$ is closed and convex
- E is convex for arbitrary $\phi(x, y) > 0$ on the unit circle
- E is closed if and only if $\phi(x, y) = 0$

Properties of convex functions

The behavior of convex function at the boundary of their domain can be disappointing, but their behavior in the interior of its domain is very simple.

Locally Bounded

Let E be convex and $u \in \mathrm{int}\,\mathrm{dom}(E)$, then E is locally upper bounded at u.

Continuity of Convex Functions

If $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is convex, then E is locally Lipschitz (and hence continuous) on int(dom(E)).

Proofs: Board.

Existence of minimizers

Theorem: Existence of minimizers

Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be l.s.c. and let there exist an α such that the sublevelset

$$\{u \in \mathbb{R}^n \mid E(u) \leq \alpha\}$$

is nonempty and bounded, then there exists

$$\hat{u} \in \arg\min_u E(u)$$

Proof: Board.

Theorem: Equivalence of l.s.c. and closedness

For $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ the following two statements are equivalent

- E is lower semi-continuous (l.s.c.)
- E is closed (its epigraph is closed)

Proof: Board.

Existence and uniqueness of minimizers

Definition: Coercivity

A function $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is called *coercive* if $E(v_n) \to \infty$ for all sequences $(v_n)_n$ with $\|v_n\| \to \infty$.

Remark: Coercivity implies existence of a bounded sublevelset

Existence of a minimizer for function with full domain

Let $E : \mathbb{R}^n \to \mathbb{R}$ be convex and coercive, then an element $\hat{u} \in \arg\min_u E(u)$ exists.

Proof:

- $\text{dom}(E) = \mathbb{R}^n$, E convex $\Rightarrow$ E is continuous.
- E is coercive, i.e. there exists a non-empty bounded sublevelset.

Theorem: Uniqueness

If $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ is strictly convex, then there exists at most one local minimum which is the unique global minimum.

Optimality conditions: differentiable unconstrained problem

How can we determine if

$$\hat{u} \in \arg\min_{u \in \mathbb{R}^n} E(u)? \tag{1}$$

Corollary to Monotonicity of the Gradient

Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, convex, and differentiable at $u \in \text{dom}(E)$. If $\nabla E(u) = 0$ then u is a global minimum of E.

Examples: derive the optimality conditions for

$$E(u) = \|u - f\|_2^2 = \sum_{i=1}^{n} (u_i - f_i)^2$$

$$E(u) = \|Au - f\|_2^2 \text{ for a matrix } A \in \mathbb{R}^{m \times n}$$

$$E(u) = \|u\|_1 = \sum_{i=1}^{n} |u_i|$$

We need a theory for non-differentiable functions (like ℓ^1)

The subdifferential

Definition: Subdifferential

Let $E : \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex, the subdifferential of E at u is

$$\partial E(u) = \{p \in \mathbb{R}^n \mid E(v) - E(u) - \langle p, v - u\rangle \geq 0, \ \forall v \in \mathbb{R}^n\}$$

- Elements of $\partial E(u)$ are called subgradients.
- If $\partial E(u) \neq \emptyset$, we call E subdifferentiable at u.
- By convention, $\partial E(u) = \emptyset$ for $u \notin \mathrm{dom}(E)$.

Example: $E(x) = |x|$ has $\partial E(0) = [-1, 1]$

$$\forall g \in [-1, 1], \quad E(x) = |x| \geq gx = E(0) + g(x - 0)$$

$\partial E(u)$ is closed and convex because it is defined by a set of linear constraints.

Geometric Interpretation

Definition: Supporting Hyperplane

A supporting hyperplane to a set $S \subset \mathbb{R}^n$ is a hyperplane $\{x \in \mathbb{R}^n \mid \langle a, x\rangle = b\}$, $a \neq 0$, such that

- $S \subset \{x \in \mathbb{R}^n \mid \langle a, x\rangle \leq b\}$ or $S \subset \{x \in \mathbb{R}^n \mid \langle a, x\rangle \geq b\}$
- $\exists y \in \partial S$ (the boundary of S) such that $\langle a, y\rangle = b$.

Theorem

Any subgradient $p \in \partial E(u)$ represents a non-vertical supporting hyperplane to epi(E) at $(u, E(u))$

Let $p \in \partial E(u)$. Then

$$
\begin{aligned}
& E(v) - E(u) - \langle p, v-u\rangle \geq 0 && \forall v \in \mathbb{R}^n \\
\Rightarrow\quad & \alpha - E(u) - \langle p, v-u\rangle \geq 0 && \forall (v,\alpha) \in \text{epi}(E) \\
\Rightarrow\quad & \left\langle \begin{bmatrix} -p \\ 1 \end{bmatrix}, \begin{bmatrix} v \\ \alpha \end{bmatrix} - \begin{bmatrix} u \\ E(u) \end{bmatrix} \right\rangle \geq 0 && \forall (v,\alpha) \in \text{epi}(E).
\end{aligned}
$$

Subdifferential Properties

Theorem: Optimality condition

Let $0 \in \partial E(\hat{u})$, then $\hat{u} \in \arg\min_u E(u)$

Proof: immediate from definition of subgradient

Subdifferential and derivatives

Let the convex function $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be differentiable at $u \in \text{int}(\text{dom}(E))$. Then

$$\partial E(u) = \{\nabla E(u)\}.$$

Subdifferentiability and Convexity

The subdifferentiability of a function implies its convexity.

Theorem

If for any $u \in \mathrm{dom}(E)$ the subdifferential $\partial E(u)$ is non-empty, then E is a convex function.

The converse statement is also true.

Theorem: Nesterov, Th. 3.1.1

If E is a closed convex function and $u \in \mathrm{int}(\mathrm{dom}(E))$, then $\partial E(u)$ is a non-empty bounded set.

The conditions of this theorem cannot be relaxed, e.g., $E(u) = -\sqrt{u}$ is convex and closed in its domain $\{u\colon u \geq 0\}$, but its subdifferential does not exists at 0.

Subdifferential rules

Theorem: Sum rule (Nesterov, Lemma. 3.1.9)

Let E_1, E_2 be convex functions such that

$$\text{int}(\text{dom}(E_1)) \cap \text{int}(\text{dom}(E_2)) \neq \emptyset,$$

then $\partial(E_1 + E_2)(u) = \partial E_1(u) + \partial E_2(u)$

Theorem: Chain rule (Nesterov, Lemma. 3.1.8)

If $A \in \mathbb{R}^{m \times n}$, $E : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ is convex, and $\text{int}(\text{dom}(E)) \cap \text{range}(A) \neq \emptyset$, then $\partial(E \circ A)(u) = A^* \partial E(Au)$

Examples: compute $\partial E(u)$

$$\begin{aligned} E(u) &= |u| \\ E(u) &= \sum_{i=1}^{m} | < a_i, u > -b_i| \\ E(u) &= \|u\|_1 \end{aligned} \tag{2}$$

Summary

- **Convex functions**
  - Every local minimum is global
  - First order optimality condition is sufficient
- The **optimality condition** for $\hat{u}$ to minimize E is

$$0 \in \partial E(\hat{u})$$

- **The subdifferential** $\partial E(u)$
  - is set valued.
  - generalizes the derivative.
  - $\partial E(u) = \{\nabla E(u)\}$ is E is differentiable at u.
  - can be identified with supporting hyperplanes to epi(E).
  - Obeys the “usual” sum and chain rules.

We now have all tools that are necessary to discuss a first class of minimization algorithms for determining

$$\hat{u} \in \underset{u}{\operatorname{argmin}}\, E(u)$$

Up next: Gradient-based Algorithms.